

1/4
ARL 71-0321

DECEMBER 1971



Aerospace Research Laboratories

ON WEAK CONVERGENCE OF EMPIRICAL PROCESSES FOR RANDOM NUMBER OF INDEPENDENT STOCHASTIC VECTORS

PRANAB KUMAR SEN

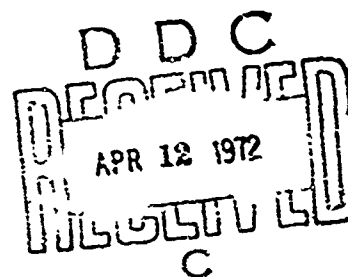
**UNIVERSITY OF NORTH CAROLINA
CHAPEL HILL, NORTH CAROLINA**

CONTRACT NO. F33615-71-C-1927

PROJECT NO. 7071

Reproduced by
**NATIONAL TECHNICAL
INFORMATION SERVICE**
Springfield, Va 22151

Approved for public release; distribution unlimited.



AIR FORCE SYSTEMS COMMAND

United States Air Force

18

NOTICES

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely related Government procurement operation, the United States Government thereby incurs no responsibility nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

Agencies of the Department of Defense, qualified contractors and other government agencies may obtain copies from the

Defense Documentation Center
Cameron Station
Alexandria, Virginia 22314

This document has been released to the

CLEARINGHOUSE
U. S. Department of Commerce
Springfield, Virginia 22151

for the sale to the public.

REPRODUCTION OF <input type="checkbox"/> COPY <input checked="" type="checkbox"/> WHITE SECTION 12 <input type="checkbox"/> NEG <input type="checkbox"/> DIFF SECTION <input type="checkbox"/> <input type="checkbox"/> UNCLASSIFIED <input type="checkbox"/>	
JUSTIFICATION: Copies of ARL Technical Documentary Reports should not be returned to Aerospace Research Laboratories unless return is required by security considerations, contractual obligations or notices on a specified document.	
BY: _____ DISTRIBUTION/AVAILABILITY STATEMENT: AIR FORCE 6-172/520	
DISC.	ATAIL. and/or SPECIAL
<div style="font-size: 2em; font-weight: bold;">A</div>	

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R & D		
(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)		
1. ORIGINATING ACTIVITY (Corporate author)		2a. REPORT SECURITY CLASSIFICATION
University of North Carolina Department of Biostatistics Chapel Hill, North Carolina 27514		Unclassified
		2b. GROUP
		M-2
3. REPORT TITLE		
On Weak Convergence of Empirical Processes for Random Number of Independent Stochastic Vectors		
4. DESCRIPTIVE NOTES (Type of report and inclusive dates)		
Scientific Interim		
5. AUTHOR(S) (First name, middle initial, last name)		
Pranab Kumar Sen		
6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
December 1971	15	7
8a. CONTRACT OR GRANT NO.		8b. ORIGINATOR'S REPORT NUMBER(S)
F 33615-71-C-1927		
a. PROJECT NO. 7071-00-12		
c. DoD Element 61102 F		8b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d. DoD Subelement 681304		ARL 71-0321
10. DISTRIBUTION STATEMENT		
Approved for public release; distribution unlimited.		
11. SUPPLEMENTARY NOTES		12. SPONSORING MILITARY ACTIVITY
TECH OTHER		Aerospace Research Laboratories (LB) Wright-Patterson AFB, Ohio 45433
13. ABSTRACT		
<p>By the use of a semi-martingale property of the Kolmogorov supremum, the results of Pyke [<u>Proc. Cambridge Phil. Soc.</u> 64 (1968), 155-160] on the weak convergence of the empirical process with random sample size are simplified and extended to the case of $p(>1)$-dimensional stochastic vectors.</p>		

DD FORM 1473

Unclassified

Security Classification

Unclassified

Security Classification

14	KEY WORDS	LINK A		LINK B		LINK C	
		ROLE	WT	ROLE	WT	ROLE	WT
	Weak Convergence Empirical Processes Stochastic Vectors Gaussian Processes						

Unclassified

Security Classification

ARL 71-0321

**ON WEAK CONVERGENCE OF EMPIRICAL
PROCESSES FOR RANDOM NUMBER OF
INDEPENDENT STOCHASTIC VECTORS**

PRANAB KUMAR SEN
UNIVERSITY OF NORTH CAROLINA
CHAPEL HILL, NORTH CAROLINA

DECEMBER 1971

CONTRACT F33615-71-C-1927
PROJECT 7071

Approved for public release; distribution unlimited.

AEROSPACE RESEARCH LABORATORIES
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
WRIGHT-PATTERSON AIR FORCE BASE, OHIO

FOREWORD

This is an interim report of the work done under Contract F 33615-71-C-1927 by the University of North Carolina. The work done by Pranab Kumar Sen in this report is sponsored by the Aerospace Research Laboratories under the above contract; it was accomplished on Project 7071, "Research in Applied Mathematics" and is technically monitored by P. R. Krishnaiah of the Aerospace Research Laboratories.

ABSTRACT

By the use of a semi-martingale property of the Kolmogorov supremum, the results of Pyke [Proc. Cambridge Phil. Soc. 64 (1968), 155-160] on the weak convergence of the empirical process with random sample size are simplified and extended to the case of $p(>1)$ -dimensional stochastic vectors.

TABLE OF CONTENTS

Section	Pages
1. Introduction	1
2. Some Preliminary Results	2
3. The Proof of Theorem 1	7
References	10

1. Introduction. Consider a sequence $\{X_i = (X_{i1}, \dots, X_{ip})', i \geq 1\}$ of independent and identically distributed stochastic $p(>1)$ -vectors, defined on a probability space (Ω, \mathcal{A}, P) , with each X_i having a continuous distribution function (df) $F(x)$, $x \in R^p$, the p -dimensional Euclidean space. We denote the marginal df of X_{ij} by $F_{[j]}$, let $Y_{ij} = F_{[j]}(X_{ij})$, $j=1, \dots, p$, $Y_i = (Y_{i1}, \dots, Y_{ip})'$, $i \geq 1$, $\underline{t} = (t_1, \dots, t_p)'$, and define

$$G(\underline{t}) = P\{Y_{ij} \leq t_j, j=1, \dots, p\}, \underline{t} \in E^p,$$

where $E^p = \{\underline{t}: 0 \leq t_j \leq 1, j=1, \dots, p\}$. Then, the empirical df for Y_1, \dots, Y_n is defined by

$$(1.2) \quad G_n(\underline{t}) = n^{-1} \sum_{i=1}^n c(\underline{t} - Y_i), \underline{t} \in E^p,$$

where $c(\underline{u}) = 1$ iff $u_j \geq 0$, $j=1, \dots, p$; otherwise, $c(\underline{u}) = 0$. Consider then the empirical process

$$(1.3) \quad W_n(\underline{t}) = n^{1/2}[G_n(\underline{t}) - G(\underline{t})], \underline{t} \in E^p,$$

and denote by

$$(1.4) \quad W_n = \{W_n(\underline{t}): \underline{t} \in E^p\}.$$

For $p=1$, it is well-known that W_n weakly converges to a Brownian motion $W^0 = \{W^0(t): 0 \leq t \leq 1\}$. For $p \geq 1$, on the space $D^p[0,1]$ of all real functions on E^p with no discontinuities of the second kind, W_n converges in distribution (in the (extended) Skorokhod J_1 -topology) to an appropriate Gaussian function,

say, $W = \{W(\underline{t}) : \underline{t} \in E^p\}$, where $E[W(\underline{t})] = 0$, and

$$(1.5) \quad E[W(\underline{s})W(\underline{t})] = G(\underline{t} \wedge \underline{s}) - G(\underline{t})G(\underline{s}), \quad \underline{t}, \underline{s} \in E^p,$$

and $\underline{t} \wedge \underline{s} = (t_1 \wedge s_1, \dots, t_p \wedge s_p)$, where $a \wedge b = \min(a, b)$; we refer to Neuhaus (1971) who also reviews the earlier literature.

Let now $\{N_v, v \geq 1\}$ be a sequence of positive integer-valued random variables, such that

$$(1.6) \quad v^{-1}N_v \rightarrow \xi, \text{ in probability, as } v \rightarrow \infty,$$

where ξ is a positive random variable defined on the same probability space (Ω, \mathcal{A}, P) .

For $p=1$, Pyke (1968) has shown that under (1.6), W_{N_v} converges in law to W^0 ; his result is extended here to the general multivariate case.

Theorem 1. Under (1.6), for every $p \geq 1$,

$$W_{N_v} \xrightarrow{\mathcal{L}} W, \text{ in the Skorokhod } J_1\text{-typology on } D^p[0,1].$$

The proof is outlined in section 3. Whereas, Pyke's arguments rely heavily on the properties of an equivalently defined Poisson process (which may become quite complicated for $p \geq 1$), our approach is based on a simple semi-martingale property of the Kolmogorov supremum, which is considered first in section 2.

2. Some preliminary results. For two real valued functions $Z(\underline{t})$ and $Z^*(\underline{t})$, defined on E^p , we let

$$(2.1) \quad \rho(Z, Z^*) = \sup\{|Z(\underline{t}) - Z^*(\underline{t})| : \underline{t} \in E^p\},$$

and for every $n \geq 1$, let

$$(2.2) \quad W_n^+ = \sup\{W_n(t) : t \in E^P\}, \quad W_n^- = \sup\{-W_n(t) : t \in E^P\};$$

$$(2.3) \quad W_n^* = \max\{W_n^+, W_n^-\} = \rho(W_n, 0).$$

Let \mathcal{F}_n be the σ -field generated by $\{X_1, \dots, X_n\}$, so that \mathcal{F}_n is \uparrow in n ($n \geq 1$).

Then, we have the following

Lemma 2.1. $\{n^{1/2}W_n^+, \mathcal{F}_n; n \geq 1\}$ and $\{n^{1/2}W_n^-, \mathcal{F}_n; n \geq 1\}$ are both non-negative semi-martingale sequences.

Proof. We only prove the result for W_n^+ , as the other follows similarly. Note that W_n^+ is, by definition, non-negative [as $W_n(t) \geq 0$ for $t=0$ or $t=1$]. Let $t_n^0 \in E^P$ be a point such that

$$(2.4) \quad W_n^+ = W_n(t_n^0); \quad t_n^0 \text{ need not be unique.}$$

Then, by definition,

$$(2.5) \quad (n+1)^{1/2}W_{n+1}^+ = (n+1)^{1/2} \sup_{t \in E^P} W_{n+1}(t) \geq (n+1)^{1/2}W_{n+1}(t_n^0),$$

so that, by (1.2), (1.3) and (2.5), for every $n \geq 1$,

$$\begin{aligned} (2.6) \quad E\{(n+1)^{1/2}W_{n+1}^+ | \mathcal{F}_n\} &\geq E\{(n+1)^{1/2}W_{n+1}(t_n^0) | \mathcal{F}_n\} \\ &= \sum_{i=1}^{n+1} E\{[c(t_n^0 - Y_i) - G(t_n^0)] | \mathcal{F}_n\} \\ &= \sum_{i=1}^n [c(t_n^0 - Y_i) - G(t_n^0)] + E\{[c(t_n^0 - Y_{n+1}) - G(t_n^0)] | \mathcal{F}_n\} \\ &= n^{1/2}W_n^+ + 0 = n^{1/2}W_n^+, \end{aligned}$$

as, given \mathcal{F}_n , $c(t_n^0 - Y_{n+1})$ assumes the values 1 and 0 with respective conditional probabilities $G(t_n^0)$ and $1-G(t_n^0)$. Q.E.D.

Lemma 2.2. For every $n \geq 1$, there exist two positive constants c_0 and c_1 , independent of n , such that

$$(2.7) \quad E\{(W_n^+)^2\} \leq c_0/c_1 \quad \text{and} \quad E\{(W_n^-)^2\} \leq c_0/c_1.$$

Proof. By partial integration,

$$(2.8) \quad E\{(W_n^+)^2\} = 2 \int_0^\infty x P\{W_n^+ > x\} dx,$$

where by Theorem 1 of Kiefer and Wolfowitz (1958), for all $n \geq 1$,

$$(2.9) \quad P\{W_n^+ > x\} < c_0 \exp\{-c_1 x^2\} \text{ for all } x \geq 0.$$

Consequently, by (2.8) and (2.9), $E\{(W_n^+)^2\} \leq c_0/c_1$. The other result follows similarly.

Lemma 2.3. For every $\epsilon > 0$, there exists a positive $K_\epsilon (< \infty)$, such that for every $n \geq 1$,

$$(2.10) \quad P\{\max_{1 \leq k \leq n} (k/n)^{1/2} \rho(W_k, 0) > K_\epsilon\} < \epsilon.$$

Proof. By (2.3), for every $\epsilon > 0$,

$$(2.11) \quad \begin{aligned} & P\{\max_{1 \leq k \leq n} (k/n)^{1/2} \rho(W_k, 0) > K_\epsilon\} \\ & \leq P\{\max_{1 \leq k \leq n} k^{1/2} W_k^+ > \frac{1}{2} K_\epsilon\} + P\{\max_{1 \leq k \leq n} k^{1/2} W_k^- > \frac{1}{2} K_\epsilon\}, \end{aligned}$$

and hence, by Lemma 2.1 along with the Kolmogorov inequality for semi-martingales

[viz., Feller (1966, p. 235)], the right hand side of (2.11) is bounded above by

$$\begin{aligned}
 (2.12) \quad & (nK_\epsilon^2)^{-1} [nE\{(W_n^+)^2\} + nE\{(W_n^-)^2\}] \\
 &= [E\{(W_n^+)^2\} + E\{(W_n^-)^2\}]/K_\epsilon^2 \\
 &\leq 2c_0/c_1 K_\epsilon^2, \text{ by Lemma 2.2.}
 \end{aligned}$$

The proof then follows by selecting $K_\epsilon > [2c_0/c_1 \epsilon]^{1/2}$. Q.E.D.

Lemma 2.4. (Uniform continuity in probability). For every $\epsilon > 0$ and $\eta > 0$, there exists a $\delta > 0$ and an $n_0(\epsilon, \eta)$, such that for $n > n_0(\epsilon, \eta)$,

$$(2.13) \quad P\left\{k: \max_{|k-n| < \delta n} \rho(W_k, W_n) > \epsilon\right\} < \eta.$$

Proof. Proceeding as in the proof of Theorem 2.1 of Pyke (1968), namely, as in his (2.7) through (2.10), we are only to show that as $n \rightarrow \infty$,

$$(2.14) \quad \max_{1 \leq k \leq n} (k/n)^{1/2} \rho(W_k, 0) = O_p(1),$$

$$(2.15) \quad \rho(W_n, 0) = \sup\{|W_n(t)|: t \in E^D\} = O_p(1).$$

Now, (2.14) has already been proved in Lemma 2.3, while by Theorem 3.1 of Neuhaus (1971) along with his treatment on the weak convergence of W_n to W , it follows that for every $\epsilon > 0$, there exists a positive $M_\epsilon(<\infty)$, such that

$$\begin{aligned}
 (2.16) \quad & \lim_{n \rightarrow \infty} P\{\rho(W_n, 0) > M_\epsilon\} \\
 &= P\{\rho(W, 0) > M_\epsilon\} < \epsilon'; \quad 0 < \epsilon' < \epsilon,
 \end{aligned}$$

which completes the proof of the lemma.

We now show that $\{W_n\}$ is a mixing sequence in the sense of Rényi (1958). This follows by defining

$$(2.17) \quad W'_n(t) = n^{-1} \left\{ \sum_{i=k_n}^n [c(t-Y_i) - G(t)] \right\}, \quad t \in \mathbb{R}^p,$$

where $k_n \rightarrow \infty$ but $n^{-1/2} k_n \rightarrow 0$ as $n \rightarrow \infty$, and noting that

$$(2.18) \quad \rho(W_n, W'_n) \leq n^{-1/2} k_n \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Consequently, proceeding as in the proof of Lemma 3 of Blum, Hanson and Rosenblatt (1963), we obtain from Lemma 2.4, the following.

Lemma 2.5. If $A \in \mathcal{A}$, then for every $\epsilon > 0$ and $\eta > 0$, there exists a $\delta > 0$, such that as $n \rightarrow \infty$,

$$(2.19) \quad P\left\{ \max_{k: |k-n| < \delta n} \rho(W_k, W_n) > \epsilon \mid A \right\} < \eta.$$

Let us now define

$$(2.20) \quad \omega_\delta(W_n) = \sup\{|W_n(t) - W_n(t')| : |t - t'| < \delta\}.$$

Then, from the results of section 5 of Neuhaus (1971), for every $\epsilon > 0$ and $\eta > 0$, there exists a $\delta > 0$, such that

$$(2.21) \quad \lim_{n \rightarrow \infty} P\{\omega_\delta(W_n) > \epsilon\} < \eta.$$

Hence, again using (2.18) and Rényi's (1958) idea of mixing sequence of sets, we have for $A \in \mathcal{A}$,

$$(2.22) \quad \lim_{n \rightarrow \infty} P\{\omega_\delta(W_n) > \epsilon \mid A\} < \eta.$$

3. The proof of Theorem 1. Let $[s]$ denote the largest integer $\leq s$. Then for $\epsilon > 0$,

$$\begin{aligned}
 (3.1) \quad & P\{\rho(W_{N_V}, W_{[v\xi]}) > \epsilon\} \\
 & \leq P\{|v^{-1}N_V - \delta| \geq \delta'\} + P\{\rho(W_{N_V}, W_{[v\xi]}) > \epsilon, |v^{-1}N_V - \delta| < \delta'\} \\
 & \leq P\{|v^{-1}N_V - \xi| \geq \delta'\} + P\{k: |k - [v\xi]| < \delta'v, \max_{k: |k - [v\xi]| < \delta'v} \rho(W_k, W_{[v\xi]}) > \epsilon\}, \quad \delta' > 0.
 \end{aligned}$$

Thus, if $\xi = c$, a positive constant, with probability one [the case treated in Pyke (1968)], it readily follows from (1.6) and (2.19) that the right hand side of (3.1) can be bounded by $\eta(>0)$ by a proper choice of $\delta' > 0$. The proof of the theorem then follows by noting that by the results of Neuhaus (1971), as $v \rightarrow \infty$,

$$(3.2) \quad W_{[vc]} \xrightarrow{D} W, \text{ in the Skorokhod } J_1\text{-typology on } D^P[0,1].$$

So, in the sequel, we consider the general case of ξ having an arbitrary distribution on $(0, \infty)$. For every $\eta > 0$, there exists an $a_0 = a_0(\eta)$, such that

$$(3.3) \quad P\{\xi \leq a_0(\eta)\} < \frac{1}{4} \eta.$$

Consider then a countable set of events

$$(3.4) \quad A_h = \{\xi: a_0(\eta) + h\delta' < \xi \leq a_0(\eta) + (h+1)\delta'\}, \quad h=0,1,\dots,$$

and let $a_h = a_h(\delta', \eta) = a_0(\eta) + (h + \frac{1}{2})\delta'$, $h=0,1,\dots$. Then, the right hand side of (3.1) is bounded above by

$$\begin{aligned}
 (3.5) \quad & P\{|v^{-1}N_v - \xi| \geq \delta'\} + P\{\xi \leq a_0(n)\} + \\
 & \sum_{h=0}^{\infty} P\{k: \max_{|k - [v\xi]| < v\delta'} \rho(W_k, W_{[v\xi]}) > \varepsilon | A_h\} P(A_h) \\
 & \leq P\{|v^{-1}N_v - \xi| \geq \delta'\} + P\{\xi \leq a_0(n)\} + \\
 & \sum_{h=1}^{\infty} P\{k: \max_{|k - [va_h]| < \frac{3}{2}\delta'v} \rho(W_k, W_{[va_h]}) > \varepsilon | A_h\} P(A_h).
 \end{aligned}$$

Now, by (1.6) and (3.3), the first two terms on the right hand side of (3.5) are bounded by $\eta/4$ by proper choice of $\delta' > 0$, while by (2.19), the last term can also be bounded by $\eta/2$, by proper choice of $\delta' (> 0)$, as $va_h \rightarrow \infty$ with $v \rightarrow \infty$, for every $h > 0$. Consequently, as $v \rightarrow \infty$,

$$(3.6) \quad \rho(W_N, W_{[v\xi]}) \xrightarrow{P} 0.$$

Thus, it suffices to show that as $v \rightarrow \infty$,

$$(3.7) \quad W_{[v\xi]} \xrightarrow{D} W, \text{ in the Skorokhod } J_1\text{-topology on } D^P[0,1].$$

Now, (3.2) implies the convergence of the finite dimensional distributions of $\{W_v\}$ to those of W , while (2.15) implies that for any $t \in E^D$, $\{W_v(t): |v-n| < \delta n\}$ satisfy the "uniform continuity in probability" condition; these two conditions, in accordance with Theorem 1 of Mogorodi (1965), imply the convergence of the finite dimensional distributions of $\{W_{[v\xi]}\}$ to those of W . So, to complete the proof of the theorem, we require to establish the 'tightness' property of $\{W_{[v\xi]}\}$ when $v \rightarrow \infty$. By (1.7) and (3.5) of Neuhaus (1971), it suffices to show that for every $\varepsilon > 0$ and $\eta > 0$, there exists a positive δ , such that as $v \rightarrow \infty$,

$$(3.8) \quad P\{\omega_\delta(W_{[v\xi]}) \geq \varepsilon\} < \eta.$$

To show this, we note that for every $\varepsilon > 0$, $\delta' > 0$,

$$\begin{aligned}
 (3.9) \quad & P\{\omega_\delta, (W_{[v\xi]}) \geq \varepsilon\} \\
 & \leq P\{\xi \leq a_0(n)\} + P\{\omega_\delta, (W_{[v\xi]}) \geq \varepsilon, \xi > a_0(n)\} \\
 & = P\{\xi \leq a_0(n)\} + \sum_{h=0}^{\infty} P\{\omega_\delta, (W_{[v\xi]}) \geq \varepsilon | A_h\} P(A_h) \\
 & \leq P\{\xi \leq a_0(n)\} + \sum_{h=0}^{\infty} P\{\omega_\delta, (W_{[va_h]}) \geq \frac{1}{3}\varepsilon | A_h\} P(A_h) \\
 & \quad + \sum_{h=0}^{\infty} P\{0(W_{[v\xi]}, W_{[va_h]}) \geq \frac{1}{3}\varepsilon | A_h\} P(A_h) \\
 & \leq P\{\xi \leq a_0(n)\} + \sum_{h=0}^{\infty} P\{\omega_\delta, (W_{[va_h]}) \geq \frac{2}{3}\varepsilon | A_h\} P(A_h) \\
 & \quad + \sum_{h=0}^{\infty} P\{k: |k - [va_h]| < \frac{1}{2}\delta', v \in (W_k, W_{[va_h]}) \geq \frac{1}{3}\varepsilon | A_h\} P(A_h),
 \end{aligned}$$

which, by (3.3), (2.22) and (2.19), can be made smaller than $\eta(>0)$ by a proper choice of $\delta'(>0)$. Q.E.D.

REFERENCES

- [1] BLUM, J. R., HANSON, D. L., AND ROSENBLATT, J. I. (1963). On the central limit theorem for the sum of a random number of independent random variables. Zeit. Wahrsch Verw. Geb. 1, 389-393.
- [2] FELLER, W. (1966). An introduction to Probability Theory and its Applications, Vol 2. John Wiley, New York.
- [3] KIEFER, J., AND WOLFOWITZ, J. (1958). On the deviations of the empirical distribution function of vector chance variables. Trans. Amer. Math. Soc. 87, 173-186.
- [4] HOGYORODI, J. (1965). Limit distributions for sequences of random variables with random indices. Trans. 4th Prague Confer. Infor. Th. Statist. Dec. fn. Random. Proc. 463-470.
- [5] NEUHAUS, C. (1971). On weak convergence of stochastic processes with multidimensional time parameter. Ann. Math. Statist. 42, 1285-1295.
- [6] PYKE, R. (1968). The weak convergence of the empirical process with random sample size. Proc. Cambridge Phil. Soc. 64, 155-160.
- [7] RÉNYI, A. (1958). On mixing sequences of sets. Acta Math. Acad. Sc. Hungar. 9, 215-228.